



UNIVERSITI PUTRA MALAYSIA

**CREATION OF A POTENTIAL GENE CASSETTE FOR RECOMBINANT
PROTEIN EXPRESSION IN GREEN MICROALGAL SYSTEM**

SAEID KADKHODAEI

FBSB 2014 8



**CREATION OF A POTENTIAL GENE CASSETTE FOR RECOMBINANT
PROTEIN EXPRESSION IN GREEN MICROALGAL SYSTEM**

By

SAEID KADKHODAEI

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in
Fulfillment of the Requirement for the Degree of Doctor of Philosophy**

July, 2014

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs, and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright© Universiti Putra Malaysia

DEDICATION

TO YOU ...!

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

**CREATION OF A POTENTIAL GENE CASSETTE FOR RECOMBINANT
PROTEIN EXPRESSION IN GREEN MICROALGAL SYSTEM**

By

SAEID KADKHODAEI

July 2014

Chairman: Professor Arbakariya B. Ariff, PhD

Faculty: Biotechnology and Biomolecular Sciences

Development of efficient strategies and reliable systems for production of recombinant proteins in high quality and quantity to reach the market are gaining more attraction. On the economic side, improvement of the efficiency and cost effectiveness of production process is highly required to be commercially viable. Although many strategies have been theoretically established to increase the level of gene expression in different specific host systems, it is needed to be practically applied. Due to the lack of enough information for expression of recombinant proteins in microalgal systems, the present study aimed at creation of a gene cassette utilizing both upstream (bioinformatics and computational analysis, genetic engineering and molecular biology) and downstream factors (bioprocess engineering) which may have critical roles in gene expression with the emphasis on two microalgae species, *Dunaliella salina* and *Chlamydomonas reinhardtii*.

As the first step in evaluating the feasibility of molecular farming in microalgae, some of the important upstream factors including matrix attachment regions (MARs), translation initiation sites, signal peptides and gene optimization parameters were studied. An artificial universal MAR sequence was designed and developed considering the MAR specific rules and motifs. To explore the specific translation initiation site, all mRNA sequences in the microalgae *D. salina* were screened among which the non-redundant mRNAs were selected for bioinformatics analysis. A consensus sequence was finally created to be used in the specific expression vector. For the secretion systems in microalgae *D. salina*, a screening procedure was carried out to find the specific signal peptides using bioinformatics and *in silico* tools in DNA, RNA and protein databases. The predicted sequences were then selected for further characterizations and the highest

scored predicted signal was embedded in the secretion series of expression vectors. For optimization of the coding sequences, the most important parameters such as the host codon preference, RNA secondary structure, GC and CpG dinucleotides content were taken into consideration. The final expression cassette containing all of the required elements including MAR, signal peptide, translation initiation site, KDEL retention signal, His 6x purification tag, V5 epitope and protease cleavage site was integrated into 25 expression vectors (25 vectors containing different elements). To develop a proprietary series of microalgae specific expression vectors, nine fragments containing the Gateway recombination cassette, optimized genes of interest (2), promoters (2), 5'-UTR, interon, 3'-UTR (2), MARs (2) and pUC18 backbone were amplified using the most high fidelity enzyme available (KAPA). The fragments were successfully assembled in a single recombination reaction through the novel multiple overlap extension PCR (MOE-PCR) technique developed in this study.

To assess the efficient transformation techniques for microalgae, three different methods (electroporation, glass beads and PEG-mediated) were used to transform one *Dunaliella* and three *Chlamydomonas* strains using the developed expression vectors. Glass bead method with 10 s agitation time, 4% PEG and 300 ng/ μ L plasmid DNA, as an efficient and simple technique showed the best results despite a drop in viable cells. The use of specific expression vectors resulted in more number and more stable transformants comparing a commercially available vector (pCAMBIA-3301) without the abovementioned optimized elements. The expression cassette containing optimized genetic elements could be delivered into microalgae cells and confer the resistance to the transformants for at least 3 months. For the downstream processes, the culture condition of *D. salina* was optimized to be used for future studies of the transformants. An experiment was conducted based on response surface methodology (RSM) to realize the reaction of microalgae to different phototrophic conditions including heterotrophic, mixotrophic and phototrophic. It was observed that the higher the amount of supplemented glucose up to 15 g/L, the more glucose utilization, biomass concentration and consequently protein content obtain. The results demonstrated superiority of mixotrophic conditions in constant light in improving the growth of microalgae.

This study demonstrated that the upstream factors such as MARs, translation initiation sites and gene optimization along with the optimized transformation systems and downstream processes, could be a potential procedure to be efficiently used in overexpression of recombinant protein production using microalgae as the expression platform. Considering the close relatedness of the studied microalgae and other close species, the proprietary vector series could be generally used in genetic engineering of at least green microalgae.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

**PENCIPTAAN KASET GEN BERPOTENSI UNTUK EKPRESI REKOMBINAN
PROTEIN DALAM SISTEM MICROALGAL HIJAU**

Oleh

SAEID KADKHODAEI

Julai 2014

Pengerusi: Professor Arbakariya B. Ariff, PhD

Fakulti: Biotechnology and Biomolecular Sciences

Pembangunan sesuatu strategi yang cekap dan sistem yang boleh dipercayai, khususnya untuk pengeluaran protein rekombinan yang berkualiti tinggi dan dalam kuantiti yang mencukupi untuk pasaran semakin mendapat perhatian. Dari sudut ekonomi, penambahbaikan ke atas kecekapan dan keberkesanan kos adalah sangat diperlukan untuk satu-satu proses pengeluaran itu mampu berdaya saing. Terdapat pelbagai strategi yang secara teorinya dikenalpasti dapat meningkatkan tahap ekspresi gen di dalam sistem perumah yang berbeza, yang mana secara praktikalnya ianya perlu dibangunkan khusus untuk organisma tertentu. Bagi mencapai matlamat tersebut, pelaksanaan kajian ini telah menjurus kepada penerokaan kedua-dua faktor hulu dan hilir yang mungkin mempunyai peranan kritikal dalam ekspresi gen, dengan penekanan diberikan kepada dua spesies mikroalga, *Dunaliella salina* dan *Chlamydomonas reinhardtii*. Oleh yang demikian, kajian ini telah digariskan kepada tiga bidang yang berbeza tetapi saling berkait, a) bioinformatik dan analisa perkomputeran, b) kejuruteraan genetik dan biologi molekular, dan c) kejuruteraan bioproses.

Sebagai langkah pertama dalam menilai daya saing kaedah pertanian molekul ke atas mikroalga, beberapa faktor hulu yang penting seperti kawasan pelekatan matriks (MARs), tapak permulaan translasi, peptida isyarat dan parameter-parameter pengoptimuman gen telah dikenalpasti untuk kajian lanjutan. Satu turutan MAR buatan yang universal telah direka dan dibangunkan dengan mengambilkira syarat-syarat spesifik dan motif MAR. Kualiti turutan yang direka kemudiannya disahkan menggunakan platform Genomatix SMARTest. Berkenaan tapak-tapak permulaan translasi di dalam mikroalga, oleh kerana terdapat beberapa variasi yang signifikan di

kalangan pelbagai eukariot, dan di sebaliknya pula terdapat kekurangan maklumat yang sedia ada untuk mikroalga, maka turutan ini telah cuba diterokai di dalam pangkalan-pangkalan data. Semua turutan mRNA bagi mikroalga *D. salina* di dalam NCBI disaring, yang mana mRNAs tidak-berlebihan dipilih untuk tujuan analisa bioinformatik. CLC – Bio Main Workbench telah diguna untuk menganalisa turutan-turutan berkenaan melalui penetapan kawasan penjajaran huluan dan hiliran pada codon permulaan. Akhirnya, turutan yang konsensus telah digunakan dalam pembangunan vektor. Disebabkan kurangnya maklumat eksperimen mengenai sistem rembesan di dalam mikroalga *D. salina*, satu kaji selidik telah dijalankan untuk mencari peptida isyarat yang spesifik dengan menggunakan bioinformatik dan peralatan *in silico* dalam pangkalan-pangkalan data DNA, RNA dan protein. SignalP 4.1 sebagai platform dalam talian yang paling dikenali dengan ambang yang mempunyai nilai p yang tinggi (> 0.9) telah digunapakai dalam penyaringan dataset urutan. Urutan ramalan kemudiannya telah dipilih untuk pencirian selanjutnya. Isyarat ramalan terunggul dengan skor rembesan tertinggi kemudiannya digunakan untuk berada pada bahagian hulu turutan pengkodan di dalam siri-siri rembesan vektor-vektor ekspresi untuk kajian sistem rembesan di dalam mikroalga hijau pada masa depan. Bagi tujuan mengoptimumkan turutan pengkodan, parameter yang paling penting seperti keutamaan codon perumah, struktur RNA sekunder, kandungan dinukleotida GC dan CPG, dan lain-lain telah diambil kira. Kaset ekspresi mutakhir yang mengandungi semua unsur-unsur yang diperlukan seperti MAR, peptida isyarat, tapak permulaan translasi, isyarat pengekalan KDEL, tag penulenan His 6x, epitop V5 dan tapak belahan protease diintegrasikan ke dalam vektor ekspresi (25 vektor yang mengandungi elemen-elemen yang berbeza). Untuk membangunkan satu siri proprietari vektor ekspresi yang spesifik kepada mikroalga, sembilan serpihan yang mengandungi kaset rekombinan Gateway, gen sasaran yang telah dioptimumkan (2), promoter (2), 5'-UTR, interon, 3' -UTR (2), MAR (2) dan tulang belakang pUC18 telah diampifikasi dengan menggunakan enzim yang berkesesuaian paling tinggi yang boleh diperolehi (KAPA). Serpihan-serpihan berkenaan telah dipasang melalui teknik terbaru iaitu pertindihan berganda dalam pemanjangan PCR (MOE-PCR) yang dibangunkan khusus dalam kajian ini. MOE-PCR telah berjaya menghimpunkan pelbagai fragmen secara serentak dalam satu tindakbalas rekombinan tunggal.

Dalam usaha untuk memilih dan melakukan kajian lanjutan terhadap transformants yang berjaya, satu ujian sensitiviti telah dijalankan untuk menentukan rintangan mikroalga (*D. salina* dan *C. reinhardtii*) kepada antibiotik dan racun rumpai yang berbeza, dan juga mencari mana rintangan racun-racun yang efektif. Mikroalga-mikroalga telah menunjukkan sensitiviti yang tinggi terhadap BASTA pada kepekatan sekurang-kurangnya 6 mg / L.

Untuk menilai teknik transformasi yang berkesan untuk mikroalga, tiga kaedah yang berbeza (elektroporasi, manik kaca dan pengantaraan PEG) telah digunakan untuk mengtransformasi satu strain *Dunaliella* dan tiga strain *Chlamydomonas* dengan menggunakan vektor ekspresi yang telah dibangunkan. Kaedah manik kaca sebagai teknik yang cekap dan mudah telah menunjukkan keputusan yang terbaik walaupun

terdapat penurunan dalam jumlah bilangan sel-sel yang hidup. Penggunaan vektor ekspresi yang spesifik telah menyebabkan bilangan yang tinggi dan transformants yang lebih stabil berbanding vektor diperolehi secara komersial (pCAMBIA - 3301) di mana ia tidak mengandungi unsur-unsur dioptimumkan seperti yang dinyatakan di atas. Sebagai salah satu proses hiliran, keadaan kultur *D. salina* telah dioptimumkan untuk digunapakai dalam kajian ke atas transformants yang selanjutnya. Satu eksperimen telah dijalankan berdasarkan kaedah sambutan permukaan (RSM) untuk memahami tindakbalas alga terhadap keadaan phototrophic yang berlainan, ini termasuklah heterotrofik, mixotrofik dan fototrofik. Keputusan diperolehi menunjukkan keadaan mixotrophic dengan pencahayaan tetap adalah paling sesuai dalam meningkatkan pertumbuhan mikroalga.

Kajian ini telah menunjukkan bahawa faktor-faktor huluan seperti MAR, tapak permulaan translasi, pengoptimuman gen, dan lain-lain bersama dengan sistem transformasi yang dioptimumkan, dan juga proses hiliran boleh menjadi prosedur yang berpotensi untuk digunapakai secara cekap di dalam pengeluaran protein rekombinan yang menggunakan mikroalga sebagai platform ekspresi. Memandangkan perkaitan rapat mikroalga yang dikaji dengan spesis terhampir yang lain, siri-siri vektor proprietari ini boleh diguna secara umum untuk kejuruteraan genetik yang sekurang-kurangnya melibatkan microalga hijau.

ACKNOWLEDGEMENTS

First and foremost, I wish to express my utmost thank and gratitude to Almighty Allah SWT for his blessings and giving me the ability and capacity to complete this dissertation.

I wish also to express my sincere gratitude to my supervisor, Professor Dr. Arbakariya B. Ariff, for his support.

I am also grateful to other members of my supervisory committee, Associate Professor Dr. Hamid Rajabi Memary, Professor Dr. Raha A. Rahim and Professor Dr. Ling Tau Chuan and for their supports.

I would like to thank the lecturers in UPM who taught me and improved my knowledge to conduct this study.

I would like to thank the laboratory and supporting staff at the Laboratory of Immunotherapeutics and Vaccines, Institute BioScience, Universiti Putra Malaysia for their help throughout my doctoral study.

I would also like to thank my friends, for their helps during this study. My appreciation goes to Universiti Putra Malaysia and Agricultural Biotechnology Research Institute of Iran that gave me the opportunity to accomplish my PhD.

Last but not the least, I wish to express my deepest gratitude to my family for their endless encouragements, patience and sacrifices made to finish this study.

I certify that a Thesis Examination Committee has met on 18 July 2014 to conduct the final examination of Saeid Kadkhodaei on his thesis entitled "Creation of Potential Gene Cassette for Recombinant Protein Expression in Green Microalgal System" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Shuhaimi bin Mustafa, PhD

Professor
Halal Products Research Institute
Universiti Putra Malaysia
(Chairman)

Janna Ong binti Abdullah, PhD

Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

Raja Noor Zaliha binti Raja Abd. Rahman, PhD

Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

Mahadevan Surianarayanan, PhD

Senior Lecturer
Central Leather Research Institute
India
(External Examiner)



NORITAH OMAR, PhD

Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 18 August 2014

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Arbakariya B. Ariff, PhD

Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Chairman)

Raha Abdul Rahim, PhD

Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Member)

Tau Chuan Ling, PhD

Professor

Faculty of Science

University of Malaya

(Member)

Hamid Rajabi Memari, PhD

Assistant Professor,

Biotechnology and Life Science Center and School of Agriculture,

Shahid Chamran University of Ahvaz

(Member)

BUJANG BIN KIM HUAT, PhD

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

DECLARATION

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____


Name and Matric No.: Saeid Kadkhodaei (GS24152)

Declaration by Members of Supervisory Committee

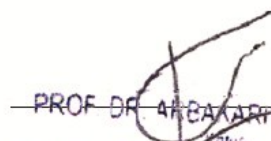
This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision,
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

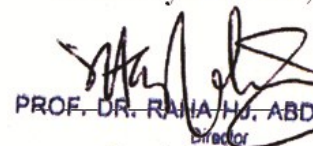
Signature:
Name of
Chairman of
Supervisory
Committee:


PROF. DR. ARBAKARIYA BIN ARIFF
Ketua
Pusat Penyelidikan Biopemprosesan dan Biopenghasilan
Fakulti Bioteknologi & Sains Biomolekul
43400 UPM Serdang, Selangor
Arbakariya B. Ariff, PhD

Signature:
Name of
Member of
Supervisory
Committee:


PROF. DR. ARBAKARIYA BIN ARIFF
Ketua
Pusat Penyelidikan Biopemprosesan dan Biopenghasilan
Fakulti Bioteknologi & Sains Biomolekul
43400 UPM Serdang, Selangor
Hamid Rajabi Memari, PhD

Signature:
Name of
Member of
Supervisory
Committee:


PROF. DR. RAHA H. ABDUL RAHIM
Director
Research Management Centre
Office of Deputy Vice Chancellor (Research & Innovation)
Universiti Putra Malaysia
Serdang, Selangor Darul Ehsan.
Raha Abdul Rahim, PhD

Signature:
Name of
Member of
Supervisory
Committee:



PROF. DR. ARBAKARIYA BIN ARIFF
Ketua
Pusat Penyelidikan Biopemprosesan dan Biopenghasilan
Fakulti Bioteknologi & Sains Biomolekul
43400 UPM Serdang, Selangor
Tau Chuan Ling, PhD

TABLE OF CONTENTS

ABSTRACT	Page i
ABSTRAK	iii
ACKNOWLEDGEMENTS	vi
APPROVAL	vii
DECLARATION	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xv
LIST OF FIGURES	xvii
 CHAPTER	
 1 INTRODUCTION	 1
 2 LITERATURE REVIEW	 4
2.1 Molecular farming	4
2.1.1 Hosts	5
2.1.2 Pros and cons in microalgae molecular farming	6
2.1.3 Advantages	7
2.1.4 Disadvantages	10
2.2 Effective factors on expression of recombinant proteins	10
2.2.1 Upstream factors	10
2.3 Microalgae	28
2.3.1 Biology	28
2.3.2 Application	29
2.3.3 Culturing Techniques	30
2.4 Concluding Remarks	31
 3 <i>IN SILICO</i> STUDIES TOWARDS HIGHER RECOMBINANT PROTEIN EXPRESSION IN MICROALGAE	 33
3.1 Introduction	33
3.1.1 Matrix attachment regions (MAR)	33
3.1.2 Translation Initiation Site (TIS)	33
3.1.3 Signal Peptides (SP)	34
3.1.4 Gene Optimization	34

3.2	Materials and methods	35
3.2.1	Development of a universal matrix attachment region (MAR) sequence	35
3.2.2	Translation Initiation Site	39
3.2.3	Signal Peptide Prediction	39
3.2.4	Gene Optimization	40
3.3	Results	41
3.3.1	Matrix Attachment Regions (MARs)	41
3.3.2	Translation Initiation Site (TIS)	47
3.3.3	Signal Peptide (SP) Prediction	50
3.3.4	Gene Optimization	55
3.4	Discussion	66
3.4.1	Matrix Attachment Regions (MARs)	66
3.4.2	Translation Initiation Site (TIS)	67
3.4.3	Signal Peptide Prediction	69
3.4.4	Gene Optimization	71
3.5	Summary	74
4	HIGH THROUGHPUT SYNTHETIC BIOLOGY BY MULTIPLE OVERLAP EXTENSION PCR	76
4.1	Introduction	76
4.2	Materials and Methods	77
4.2.1	Preparation of the Fragments	77
4.2.2	Multiple DNA Fragment Assembly (MDFA)	82
4.2.3	Sequence and Ligation Independent Cloning (SLIC)	83
4.2.4	Successive Hybridization Assembling (SHA)	83
4.2.5	Multiple overlap extension PCR (MOE-PCR):	83
4.2.6	Competent Cell Preparation	84
4.2.7	Verification Steps	84
4.3	Results	85
4.4	Discussion	91
4.5	Summary	97

5	OPTIMIZATION OF GENETIC TRANSFORMATION PROCEDURE IN THE MICROALGAE	98
5.1	Introduction	98
5.2	Material and Methods	99
5.2.1	Microalgae Species, Strains and Maintenance	99
5.2.2	Sensitivity Test and Growth Analysis	99
5.2.3	Expression Cassettes and Vectors	99
5.2.4	Transformation Techniques	100
5.2.5	Colony PCR	102
5.2.6	DNA Isolation and PCR Analysis	103
5.2.7	RNA Isolation	103
5.2.8	cDNA Synthesis and RT-PCR analysis	104
5.2.9	Statistical Analysis	104
5.3	Results	104
5.3.1	Sensitivity Test	104
5.3.2	Transformation Techniques	107
5.3.3	Colony PCR	112
5.3.4	DNA Extraction	112
5.3.5	PCR Analysis	113
5.3.6	RNA extraction and RT-PCR analysis	114
5.4	Discussion	115
5.4.1	Sensitivity Test	115
5.4.2	Colony PCR	116
5.4.3	DNA Extraction, PCR and RT-PCR analysis	117
5.4.4	Transformaton	118
5.5	Comparison of different transformation methods	120
5.6	Summary	123
6	OPTIMIZATION OF MICROPLATE-BASED GROWTH OF <i>DUNALIELLA SALINA</i> FOR PROTEIN PRODUCTION (USING RESPONSE SURFACE METHODOLOGY)	124
6.1	Introduction	124
6.2	Materials and Methods	126

6.2.1	Microalgal Species and Cultivation Procedures	126
6.2.2	Molecular Characterization of the Strains	127
6.2.3	Analytical Procedures	127
6.2.4	Experimental Design	128
6.3	Results	130
6.3.1	Molecular Characterization	130
6.3.2	Growth Profile of D ₁ , D ₂ and D ₃	130
6.3.3	Optimization of Growth Conditions Using RSM	131
6.4	Discussion	140
6.5	Summary	143
7	CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDENTS	144
7.1	Conclusions	144
7.2	Recommendations for future research	146
	REFERENCES	149
	APPENDICES	178
	BIODATA OF STUDENT	199
	LIST OF PUBLICATIONS	200